hSOLICITATION, OFFER AND AWARD *GPO08-011*	1. THIS CONTRACT IS A RATED ORDER UNDER DPAS (15 CFR 700)	RATING *Internet Archive*	PAGE 1	OF PAGES 46

2. CONTRACT NUMBER	3. SOLICITATION NUMBER GPO08-011	4. TYPE OF SOLICITATION □ SEALED BID (IFB) x NEGOTIATED (RFP)	5. DATE ISSUED	6. REQUISITION/PURCHASE NO.

7. ISSUED BY CODE

US Government Printing Office
Acquisition Services, STOP: CSAS
North Capitol and H Streets, NW
Washington, DC 20401

8. ADDRESS OFFER TO *(If other than Item 7)*

US Government Printing Office
Bid Room, C161, STOP: PPSB
North Capitol and H Streets, NW
Washington, DC 20401

NOTE: This is a negotiated (RFP)

SOLICITATION

9. Submit Proposals in accordance with Section L. Mailed and hand delivered copies shall be placed in the depository located in **Room C161, by 3:00 pm** local time, September 8, 2008.

CAUTION — LATE Submissions, Modifications, and Withdrawals: See Section L, Provision No. 52.214-7 or 52.215-1. All offers are subject to all terms and conditions contained in this solicitation.

10. FOR INFORMATION CALL:	A. NAME	B. TELEPHONE *(NO COLLECT CALLS)*			C. E-MAIL ADDRESS
		AREA CODE	NUMBER	EXT.	
	Albertha M Broadnax	202	512-0966		abroadnax@gpo.gov

11. TABLE OF CONTENTS

(✓)	SEC.	DESCRIPTION	PAGE(S)	(✓)	SEC.	DESCRIPTION	PAGE(S)
PART I – THE SCHEDULE				PART II – CONTRACT CLAUSES			
X	A	SOLICITATION/CONTRACT FORM	1	X	I	CONTRACT CLAUSES	19-28
X	B	SUPPLIES OR SERVICES AND PRICE/COST	2	PART III – LIST OF DOCUMENTS, EXHIBITS AND OTHER ATTACH.			
X	C	DESCRIPTION/SPECS./WORK STATEMENT	3-11	X	J	LIST OF ATTACHMENTS	29
X	D	PACKAGING AND MARKING	12	PART IV – REPRESENTATIONS AND INSTRUCTIONS			
X	E	INSPECTION AND ACCEPTANCE	12	X	K	REPRESENTATIONS, CERTIFICATIONS AND OTHER STATEMENTS OF OFFERORS	30-35
X	F	DELIVERIES OR PERFORMANCE	12-13				
X	G	CONTRACT ADMINISTRATION DATA	14-15	X	L	INSTRS., CONDS., AND NOTICES TO OFFERORS	36-43
X	H	SPECIAL CONTRACT REQUIREMENTS	16-18	X	M	EVALUATION FACTORS FOR AWARD	44-46

OFFER *(Must be fully completed by offeror)*

NOTE: Item 12 does not apply if the solicitation includes the provisions at 52.214-16, Minimum Bid Acceptance Period.

12 In compliance with the above, the undersigned agrees, if this offer is accepted within _____ calendar days *(60 calendar days unless a different period is inserted by the offeror)* from the date for receipt of offers specified above, to furnish any or all items upon which prices are offered at the price set opposite each item, delivered at the designated point(s), within the time specified in the schedule.

28. DISCOUNT FOR PROMPT PAYMENT *(See Section I)* ➡	10 CALENDAR DAYS %	20 CALENDAR DAYS %	30 CALENDAR DAYS %	CALENDAR DAYS %

14. ACKNOWLEDGMENT OF AMENDMENTS *(The offeror acknowledges receipt of amendments to the SOLICITATION for offerors and related documents Numbered and dated):*	AMENDMENT NO.	DATE	AMENDMENT NO.	DATE
	1	8/21/2008	2	9/11/2008

15A. NAME AND ADDRESS OF OFFEROR CODE FACILITY

Internet Archive
116 Sheridan Avenue
San Francisco, CA 94129-1711

16. NAME AND TITLE OF PERSON AUTHORIZED TO SIGN OFFER *(Type or print)*

Brewster Kahle, Digital Librarian

15B. TELEPHONE NUMBER			□ 15C. CHECK IF REMITTANCE ADDRESS IS DIFFERENT FROM ABOVE – ENTER SUCH ADDRESS IN SCHEDULE.	17. SIGNATURE	18. OFFER DATE 9/19/2008
AREA CODE 415	NUMBER 561-6767	EXT.			

AWARD *(To be completed by Government)*

19. ACCEPTED AS TO ITEMS NUMBERED	20. AMOUNT	21. ACCOUNTING AND APPROPRIATION

22. AUTHORITY FOR USING OTHER THAN FULL AND OPEN COMPETITION:	23. SUBMIT INVOICES TO ADDRESS SHOWN IN *(4 copies unless otherwise specified)* ➡	ITEM

24. ADMINISTERED BY *(If other than Item 7)* CODE	25. PAYMENT WILL BE MADE BY CODE

26. NAME OF CONTRACTING OFFICER *(Type or print)*	1. UNITED STATES OF AMERICA	28. AWARD DATE

The Internet Archive is interested in providing its digitization, hosting, access and document lifecycle management services to the Government Printing Office for digitization of its FDLP legacy collection (Solicitation GPO08-011)

Contact:
Brewster Kahle
Digital Librarian, Internet Archive
Brewster@archive.org
(415) 561-6767

FEDLINK Information:
FEDSCAN Service
Anne Harrison at (FEDLINK)
202-707-4834
anha@loc.gov

TABLE OF CONTENTS

I. EXECUTIVE SUMMARY 4

II. INTRODUCTION 6

III. TECHNICAL CAPABILITY 8

1. INTERNET ARCHIVE 8
2. LAW LIBRARY MICROFORM CONSORTIUM 22

IV. ADDITIONAL SERVICES 25

HOSTING 25
ACCESS 25

V. WORK PLAN FOR DIGITIZATION 26

VI. SPECIFIC EXPERIENCE OF KEY PERSONNEL 27

VII. FINANCING STRATEGY 29

A. LLMC MEMBERS 30
B. FDLP PARTNER LIBRARIES 30
C. THIRD-PARTY GRANTS 31

APPENDIX B: SIMILAR EXPERIENCE MATRICES 33

APPENDIX C: SCRIBE SCANNER 42

APPENDIX D: IMAGES FROM INTERNET ARCHIVE'S GOVDOCS COLLECTION 43

APPENDIX E: RESUMES OF KEY PERSONNEL 46

APPENDIX F: PARTNER LETTERS 51

I. EXECUTIVE SUMMARY

The Internet Archive (Archive) with the support of two partners proposes a five-year project to digitize and preserve the entire corpus of print Federal Depository Library Program (FDLP) legacy collection materials that are designated as in-scope by the Government Printing Office (GPO). The Archive will provide image conversion via its network of scanning centers, image post-processing, and hosting and free permanent public access to all scanned images without restrictions of any kind. The resulting digital files in tiff format will be provided to the GPO. They will also be archived and hosted for free permanent public access (PPA) by the Internet Archive, from which they will be downloadable individually and in bulk, and will be suitable both for printing and access through any search engine or website that might want to import and share copies of the documents.

The Law Library Microform Consortium (LLMC) joins the Archive in this proposal with particular focus on legal, regulatory and governance-related materials. The LLMC will take responsibility for identifying this subset of materials provided by GPO, estimated to be 26% of the legacy collection, and for generating the funds to have them scanned. In return they will keep copies of the scanned documents for access by LLMC member institutions.

The University of Florida's Dean of University Libraries, Judith Russell, also joins this proposal with the role of spearheading the funding strategy for FDLP libraries. Ms. Russell, as the former Superintendent of Documents at the GPO, understands particularly well the network of FDLP libraries and their interest in having the collections digitized. She also brings unique understanding of the FDSys requirements and technical specifications that will guide the Archive's activities under this proposal. She will coordinate interactions with the FDLP libraries for funding and potentially for sourcing materials. All three partners have identified potential sources of funding for this project with the goal of zero net cost to the GPO for digitization services. The length and scope of the project require a broad-based funding strategy that will engage diverse participants from the library, university and private sector communities. The Archive and its partners have a three-pronged strategy to raise the needed funds, including FDLP libraries, LLMC member fees, third party grants. The ultimate scope of the Internet Archive's work will be dependent upon the availability funding resulting from this strategy.

The Internet Archive has been providing mass digitization services since 2003 and operates eleven scanning centers located around the United States – several at FDLP partner libraries. These scanning centers utilize the Scribe for image capture, a non-destructive scanning station designed by the Internet Archive's technologists. The Scribe enables digitization of bound volumes at large scale without dis-binding. The Archive has the ability to scan oversized documents, foldouts and loose documents. The LLMC has the ability to scan loose and dis-bound documents as well.

The Internet Archive is able to provide multiple access formats and mechanisms including free, public searchable websites at www.archive.org, www.openlibrary.org and public.resource.org, print-on-demand capabilities optimized for the Espresso book-printing machine, and several e-book reader formats including the $100 Laptop project. Digitized files will fully meet the requirements of the FDSys Operational Specification for Converted Content.

Digitization of the GPO's legacy collection would build on a collaborative scanning project launched by the Internet Archive and public.resource.org, called the Government Documents Collection (http://www.archive.org/details/USGovernmentDocuments). This project to digitize and provide free permanent public access to government documents was launched with $500,000 in private funding and builds on 2.5 million pages already scanned. The Archive and its partners will build on this core collection and initial funding as part of its strategy to obtain funding for the entire legacy collection. The pace and extent of digitization activities over the proposed five year period will be driven both by the availability of documents from the GPO and the availability of funding to cover scanning costs.

The Internet Archive provides its digitization services through FedLink under the name FEDSCAN and has a GSA Schedule 36 application pending.

Persons authorized to negotiate on the participant's behalf with the Government in connection with this solicitation:

Brewster Kahle
Digital Librarian, Internet Archive
Phone (415) 561-6767
Brewster@archive.org
FAX (415) 840-0391

Jacques Cressaty
Director of Administration, Internet Archive
Phone: (415) 561-6767
jacques@archive.org
FAX: (415) 840-0391

II. INTRODUCTION

The Internet Archive is a 501 © 3 non-profit institution founded in 1996 with the mission of providing Universal Access to Knowledge. The Archive has been collecting, managing, digitizing and hosting digital content at a very large scale for over twelve years. It has been involved in high-volume, systematic book scanning since 2003 and continues to increase its activities in this area with partners from national and regional libraries, archives, federal agencies and numerous thematically focused digitization initiatives. The Internet Archive's digital collections now include over 500,000 scanned books and approximately 150 million pages as well as a 12 year cumulative archive of the worldwide web and collections of digitized audio recordings, films, software and courseware.

The Internet Archive designed its own open access scanning hardware, metadata management and image processing software to enable an integrated, high volume scanning and hosting workflow. The Archive's digital collections – comprising over two petabytes of content - are preserved and hosted at the Archive's Digital Repository facility in San Francisco and selected partner mirror facilities.

The Law Library Microform Consortium (LLMC) was chartered in 1976 as a non-profit 501 (c) (3) cooperative of libraries, dedicated to the twin goals of preservation of and access to legal and governance-related documents. LLMC provides its membership with a high-quality and inexpensive source of digital or film replacement for older, physically deteriorating books. LLMC also provides an economical way to complete retrospective collections or share unique documents. All content is catalogued to the highest OCLC standards.

In 2002 the 253 LLMC Participating Libraries, representing 100% of US Federal Law Libraries including the US Supreme Court and more than 91% of US law school libraries, deeded the assets and goodwill of LLMC-Fiche to a new digital service, LLMC-Digital, which is dedicated to making all the former fiche titles and more content available online. LLMC-Digital current scanning efforts resulted in more than 25,900 volumes online, representing over 1,560 titles and approximately 18 million page-images.

Ms. Russell has deep and longstanding relationships with the Federal Depository Library partners from her tenure as Superintendant of Documents at the GPO. She therefore is in a unique position to engage and harness the funding capability and document sourcing possibilities (with GPO's permission) of the FDLP libraries.

By working with the Internet Archive and its partners, the GPO can leverage an extensive scanning network and proven workflows with the additional cost-free benefits of long-term duplicate file storage and PPA for the general public. Once these materials are digitized and made available for free public access the Archive can bring to bear related initiatives in OCR correction (www.recaptcha.net) and distributed proofreading (www.distributedproofreaders.net) to improve the quality of the hosted content. These broad-based projects are currently supporting Internet Archive's complete digitized books collections.

The GPO's legacy collection would provide a model for future digitization projects by establishing unrestricted access, cost control and future software support based open source tools, and long-term preservation and maintenance based on the funding efforts of the Internet Archive, LLMC and others.

Specifically, the Internet Archive and its partners propose to provide the following services:

> **Conversion Services**: Conversion of the GPO Content to digital format with an appropriate level of markup, annotation, OCR conversion, and metadata production. Conversion will be done on Internet Archive-owned and designed non-destructive Scribe machines situated at one or more of the eleven domestic scanning locations. Images scanned at these centers will be automatically transferred to the Internet Archive's California facility, where image post-processing into required formats will be conducted. The Scribe has been designed to enable digitization of bound monographs of varying size and thickness in a non-destructive manner.
>
> Certain legal and governance-related materials may be scanned by the LLMC at their facilities. Currently LLMC does all of its step-and-repeat scanning (i.e. for bound materials) on special scanners made by a German company called SMA. This equipment has been chosen both because it produces an exceptionally good image, but also because it has a very gentle impact on the materials being scanned. For materials that can be dis-bound, LLMC currently uses high-speed scanning equipment marketed by the Canon Company. Once scanned and run through Quality Control procedures, those images will be transferred to the Internet Archive and to the GPO.
>
> Regardless of scanning location, the files will be delivered to the GPO in tiff format as specified in section B.2 of the RFP.
>
> **Access Services:** The Internet Archive will provide persistent and continuous Web-based access to the digitized GPO content. Digitized images will be archived and hosted for public access at the Internet Archive facility in California. Internet Archive technical staff will ensure optimized presentation through its web interface with integrated search capability. The LLMC will provide additional access to the legal and governance-related materials to their members through their established online channels. The ability of both of these organizations to provide free public access in addition to scanning services distinguishes them as potential partners.
>
> **Storage Services:** The Internet Archive proposes to provide permanent archiving of the digitized GPO Content and related metadata, including migration to new storage platforms, to ensure their continued functionality and availability to GPO,

FDLP partners and their user communities. This additional service, provided at no cost to the government, enhances the preservation of the digitized materials.

Scanning Centers

The Internet Archive operates scanning centers at

- Library of Congress, Washington DC
- Allen County Public Library, Ft. Wayne, IN
- UCLA, Los Angeles, CA
- San Francisco, CA
- Boston Public Library, Boston, MA
- Princeton Theological Seminary, Princeton, NJ
- University of Illinois at Urbana/Champaign, IL
- Smithsonian Institution, Washington DC
- John Hopkins University, Baltimore, MD
- University of North Carolina at Chapel Hill, NC
- North Carolina State University, ____, NC
- New York Public Library, NYC, NY

The Archive is scanning approximately 5,000 books per week, or 2+ million pages, through these centers. The distributed locations will facilitate the digitization workflow by allowing GPO to provide materials obtained from FDLP partners to the nearest scanning center. The new 10-Scribe scanning center at the Library of Congress will most likely be used for scanning documents available from the GPO headquarters. Capacity can be expanded here by an additional 10 Scribes if needed based on the pace of document flow to from the GPO. The Internet Archive may provide the option of scanning onsite at FDLP locations if it is more cost effective or logistically more beneficial to the GPO. This will only be done upon mutual agreement of the GPO and the Internet Archive.

III. TECHNICAL CAPABILITY

1. Internet Archive

a. Process for digitizing content

The Internet Archive has been digitizing bound monographs for government, academic, library and international partners since 2003. Through its eleven scanning centers the Archive has digitized over 500,000 books (approximately 100 million pages) contributed by academic institutions, public libraries, museums and other contributors. The rate of scanning has increased dramatically as new scanning centers have been brought online. The Archive scanned approximately 60 million pages in 2007 alone. Current scanning capacity is over 1,000 books per day and in excess of 9 million pages per month.

Task 1: Confirmation of receipt of materials

Retrieval, packing and shipment of Materials by GPO-

1. Materials, meeting IA specifications, will be delivered to the Internet Archive.
2. Materials will be packed on cart and wrapped for transport and shipping, unless otherwise agreed to. Any special procedures will be determined in advance.

3. A paper and a digital (excel) copy are to travel with each shipment of materials delivered to Scanning Center.

Receipt of shipment by IA prior to scanning

Upon receipt of the materials, the Archive and/or LLMC will furnish to GPO a confirmation of materials received, including number of boxes received. IA will match the count on the pick list to the materials on the cart. If count matches, books will move to the loading area. If the book count does not match- the Archive will alert the GPO.

Internet Archive and LLMC will agree with GPO upon a frequency and quantity at which materials will be sent to them, during periodic planning meetings. GPO will pay shipping costs for all materials that are shipped to the Internet Archive and/or LLMC for digitization.

The Archive understands GPO will notify them if and when no materials are available to send for digitization due to circumstances beyond the government's control, and will provide an estimated timeframe for when materials may be provided. Likewise, the Archive will notify the GPO when funding limits the ability to accept and scan new materials and will provide a forecast and estimated timeframe for accepting new materials.

Task 2: Scanning of tangible materials files

The Internet Archive will work with the GPO to assess the range of materials to be digitized and to identify any materials that would need special handling. At this point the GPO legacy documents appear to be within the scanning capabilities of the Internet Archive and/or the LLMC. Any exceptions would be identified through collaboration with GPO during the selection process, with an opportunity to develop a plan for handling them.

This section details in general, of the technical specs and the operation of how the Internet Archive's scanning equipment works.

1. Image capture-Color

2. Output formats
    a. Color images in JPEG2000 format in pixels per inch listed below.
    b. OCR in 2 XML formats: ABBYY and DJVU formats. ABBYY 6.0 is used, with its quality. As new versions and alternative vendors become available, a review will be coordinated between LP and DS before implementation. OCR XML character format is UTF-8.
    c. XML for metadata from MARC.
    d. XML for operational metadata collected during scanning.
    e. Searchable PDF.

f. XML structural metadata for monographs include page numbers when apparent on the pages that is checked by the scanner operator. These formats will be delivered from the Internet Archive servers the Internet via HTTP, FTP, RSYNC, or OAI.

3. DPI vs. Size
Example of DPI vs. size, chosen to optimally image a given size book.

DPI	**Height (inch)**	**Width (inch)**
300	14.6	9.7
400	10.9	12.5
500	8.7	6.58
600	7.3	4.9

Scanning Equipment For Bound and non-bound materials
"Scribe"

Background-
The Internet Archive has tested and evaluated many commercially available scanning devices, but felt that due to the great variety of paper types, binding types and collections to be digitized, an in-house developed Scanning solution would provide the safest and, ultimately, a cost effective way of scanning books. The equipment shown below has been field tested and has successfully scanned millions of pages with virtually no damage caused by the equipment to materials being digitized.

Non-Destructive Scanning Station –
The Scribe workstation*- which is comprised of a frame that holds two cameras on a rail, to capture both the verso and recto pages of the book to be digitized, a cradle that the book sits in (the spring supported cradle is 'v' shaped so there is minimal stress put on the book), a glass platen that is raised and lowered by means of a foot pedal to allow for the pages to be turned, two banks of lights that illuminate the book and two small computers to run the cameras and pre-process the images. The captured images are, upon the completion of the book being scanned and the QA process being completed, are uploaded via RSYNC to processing computers located in California.

- Mechanical/Electrical parameters- 7 amps per Scribe, 800 watts of heat generated per Scribe, standard UK/USA/ voltage, appx 100 sq feet per Scribe of work area, Scribe footprint is 68" long x 37" wide x 79 high; (Dimensions for shipping with out crate- 60" long-rails removed, x 32"wide-remove monitor/arm x 79in high; Dimensions for shipping with crate- 68" long x 42" wide x 86" high, 609 pounds)- NOTE check all doors for access- the weight is appx 300 lbs. For a scanning center of 10 Scribes appx 1000 sq feet of space is required.

Workflow-
Image capture

1. The Scribe scanner currently captures page images using a pair of digital single-lens-reflex (DSLR) cameras, either a 16.7 mega-pixel Canon 1DS-Mark II or a Canon EO 5D, 12.8 mega-pixel, with a Canon EF 100mm f 2.8 macro lens (http://www.usa.canon.com/app/pdf/lens/EFLensChart.pdf). IA is always evaluating new cameras and if a better solution comes along, after coordinating with GPO and DS.

2. The lighting system for illuminating the target books consists of eight (8) 5000 Kelvin, 36 degree, 35 watt museum-grade Solux bulbs, and provides a smooth daylight spectrum with a high color-rendering index.

3. Lighting compensation program- To help make the lighting even across images being scanned.

4. Reference targets- A color target (ColorChecker 24) and a white card is shot with each book for reference, which can be used for ICC-based color management.

5. Image transfer- Page images are downloaded in real time to a scanner management and image processing computer which also run the camera management software that releases the camera shutters.

6. Equipment Calibration
   1. Cameras are calibrated per Manufacturer's spec. Cameras out of spec or standard performance will be sent back to Manufacturer for repair.
   2. Lights used in scanning process are replaced as is necessary. Light comp algorithms are run daily on each unit of scanning equipment.
2. Scribe stations are calibrated and aligned before being used.

Process steps-
A. Meta Data- for all new collections that require attribution or at the beginning of a scanning center set up.

1. Meta data and set up form- includes Contributing Library, Digitization Sponsor, Collection Name, Contact details, etc. to be filled out by GPO and sent to Robert@archive.org .
2. Meta data set up is then incorporated into IA scanning/loading screens. This is to ensure proper attribution and organization of the materials is being completed.
3. If the Z39.50 set up is not being used to locate GPO catalog records, alternate confirmation of how records are to be located must be agreed to by GPO and IA. A test pick list of at least 50 records, see below, and should be generated to test that IA can locate the proper MARC records.

4. Any changes or requests by the Library that would impact the Meta data must be in writing; for example, a new collection, a new funding source or a new sub-collection.

B. Preservation and handling-

1. GPO preservation personnel will meet with IA to establish and agree on how to handle materials to be scanned, how to deal with obvious rejects, how to flag and tag materials not able to be scanned, agreement on error codes and the like. Deviations from this process will be in writing and where possible all steps are documented with a visual example for reference.
2. A review between IA/GPO on what materials can/can't be scanned is conducted, should any remain after the GPO's screening process. Questionable materials may be tested or tried before being put into the scanning plan. All selection of materials will follow these guidelines**,** See section F.
3. Rare materials are defined as materials that would normally not be in circulation or should not be included in the general scanning population.

C. Inspection of materials to be scanned by IA prior to scanning

IA, as it loads each item to be scanned, will inspect each item for possible factors that would impact it's ability to be scanned. Rejected materials will be marked in a pre-agreed format and returned with books scanned. These items may be scanned at a later date as new processes become available or a different cost structure is put in place.

D. Digitization Criteria to be used by GPO and IA-

1. Criteria to be used for determining if the materials may be digitized are listed below and will include, but not necessarily be limited to the following:

    a. Materials that have multiple titles/physical volume (eg, 63 vols, 20-40 pamphlets per vol, of "Forestry Pamphlets" without analytics) will be reviewed to ensure all proper meta data is understood.

    b. Materials will be screened for size- materials not fitting the requirements shown below will be returned unscanned.
    - 9.7 wide by 14.5 high at the max. Books as small as 3 inches x 3 inches may possibly be scanned.
    - Less than 3 inches thick are ok, greater than 3 inches will need to be reviewed.
    - Books should on average should be 200 pages or larger. If a collection is mostly under 100 pages a review should be undertaken between IA and DS to ensure the quoted price per scanned page can be met.

5. Book Style-
    - Side bound Monographs, single sheets (for LLMC), no top bound books.
    - Rebound books need to be checked for how tight the gutter or binding is, or if the text runs outside the margin (it will show the cradle).
    - When there is more than one title within a bound book; each of these has to be clearly marked with a paper strip; each of these will be counted as a separate book. Deviations from this must be approved between IA/GPO.

- Fold outs can be included as part of a special production channel.
- Soft cover books are ok if they are bound.
- Covers that are almost separated from book and appear too fragile may be rejected unless agreed to in advance by GPO.
- Materials should be of similar condition or quality to what would be put into circulation. Materials deemed not robust to go into circulation should be reviewed with Scanning Center Coordinator before scanning.
- Tight bindings that will not lay open for digitization per IA specification limits will be rejected.

6. Paper style-
    - Most paper styles can be scanned, except highly acidic paper that disintegrates to the touch. Note that if a hard-to-scan paper is to be digitized; a review of time to scan versus any 'damage' will be undertaken.
    - All pages should be pre-cut. Unless otherwise instructed, books with uncut pages will not be scanned.
    - Pages should be able to be lifted and turned with normal effort. Sticky pages or the like will not be scanned.
    - Pages should not be excessively dusty, have excessive mildew or be moldy.

7. Gutters/Margins-
    - Any book where the text is less than a quarter inch off the gutter, on an approximate 75-degree angle will be unscannable.
    - Text that runs to the edge of the page or margin can be scanned but the presentation will be poor, as the cradle will show. The GPO must approve this.

8. Bibliographic data-
    - Provided that sufficient data is submitted by GPO along with the printed materials, this should not be an issue.

9. Rejection codes to be sent back to Library will be mutually agreed upon during the project planning session. Sample codes used by the Archive are listed below:

Code	**Definition**
BI	Fragile or no binding (includes items in clam shells or phase boxes)
CAT	Cataloging error
DAM	Damaged
DAT	still in copyright
FO	Foldouts
LG	too large
MAR	margins too tight
MIS	Missing pages

MUL	multiple titles bound together
NA	not available
LAN	Outside language parameters
LIST	Picklist error
LINK	unsuccessful link to metadata
NOS	not on shelf – missing/lost
OUT	not on shelf – checked out
PAG	Pagination problems: section(s) bound out of order or upside down
PAP	Brittle paper, tissue paper
SKW	Skewed text – to point of being unreadable
SM	too small
UNC	Uncut pages (more than 5)
FOR	non-book format
VEL	Vellum
WD	Withdrawn
SPH	requires special handling
DUP	exact duplicate of another on list

E. IA loading process-

1. The book id is loaded into IA Screen to located appropriate MARC record- records not found will be cause for reject of book.
2. If IA receives a series that is cataloged under one bib record, without volume numbers, IA will either add the year as a volume number or add the volume number in IA's set up screen. This would be analogous to, in the physical world, where the series are shelved together, a person locating the book(s) from the call number and then scanning the items on the shelf for the volume they want. IA will not in a series, delete any information from the descriptions field or add to the descriptions field on the MARC record.
3. QA check is done to ensure book and MARC record match.
4. Unique identifier is created and MARC record is attached to that identifier.
5. Book is placed in queue for scanning.
6. As mentioned above, any books determined not to be scannable are set aside and a rejection form is attached.
7. Each book is also given a color-coded flag that shows the Book identifier.
8. Any special scanning instructions are included with book.

F. IA scanning process-

1. Materials to be scanned are placed in queue for scanners, typically on book carts.
2. The flag inside the material to be scanned is matched to digital file to ensure a proper match.
3. Images are scanned into appropriate digital file.
4. Images are QA'd following adjustments to ensure proper preservation or presentation.
5. Digital file is closed and uploaded to IA processing center.

G. IA processing

1. Uploaded images are processed to create storage files and access files.
2. TIFF files will be created from the scanned images, and additional files may be created including:
    a. ID.pdf
    b. ID_jp2.zip - will not be accessible, this is only for long term preservation
        i. zipped folder of the book without bookplate and watermark is specific to the sponsor, contributor; e.g. this will vary for each sponsor/library.
        ii. [ID]_nnnn.jp2 (where first image index number is the front cover and the last scan # is the back cover)
    c. ID_lib_jp2.zip
        i. zipped folder of the book with bookplate and watermark
        ii. [ID]_lib_nnnn.jp2 (where image index the first number is front cover and the last scan # is the back cover)
    d. ID_marc.xml
    e. ID_meta.mrc
    f. ID_meta.xml
    g. ID_metasource.xml
    h. ID_raw_jp2.zip , unprocessed storage format, no watermark/book plate
    i. Scandata.zip

5. Metadata will reside in meta.xml file, and will include the following required fields:
    a. Identifier
    c. Collection-
    d. Identifier-bib (unique identifier -- local catalog number; from pick list),
    e. Contributor (GPO or FDLP library)
    f. Title
    g. Volume
    h. Creator (if in MARC record)
    i. Publisher (if in MARC record)
    j. Collection (possibly multiple collection fields)
    k. Operator
    l. Scanner
    m. Scandate
    n. Identifier-access (URL for accessing this book)

6. Processing Background- The digitized image is captured initially as a camera raw file (CR2). This is run through a JPG 2000 compression to generate a raw JPG 2000 for storage. The raw JPG 2000 is then turned into a processed master which is used to generate the access formats.

    i. TIFF masters will be created for delivery to the GPO.
    ii. Access format- the processed JPG 2000 masters are compressed in a JPG 2000 format which feeds into the OCR and book generation tools. Image sizes may vary depending on the complexity of the page, but are typically

in the 760 KB range, yielding an approximate compression ratio of 20:1; relative to the camera raw image (CR2 is appx 15MB/image). PDF and DjVu; both of which are OCR'd.

iii. Quality settings will vary based on vendor tools used. For example a quality setting of 50 on a scale of 1-100 was used for the Luratech. This setting was determined based on user surveys.

H. Turnaround for processing by IA- typically 72 hours from arrival to return of book cart.

1. The goal is to derive and upload a book within 24 hours after scanning.
2. An internal IA QA step is performed inside the scanning center. Criteria for QA are outlined below in the Quality Section.
3. If the scanning lot is rejected, then the process outlined in the Quality Section __ is undertaken-
4. A scanned item is then published online within 48 hours after scanning.
5. Materials scanned are then 'curated' by IA and are available for downloading after that by the GPO.
6. Approved Materials having been scanned are then ready to be checked out and returned to the GPO.

Internet Archive understands that the GPO's technical specifications are evolving and subsequent versions of the specifications may be published. The Archive will immediately review any changes and consult with GPO on its ability to conform to the revised specifications. The Archive has every intention of conforming to such changes prospectively (not retroactively applied to previously digitized materials) within its capabilities.

Task 3: Creation of Preservation and/or Access derivative level files for GPO

As a result of the scanning process, the Archive and/or LLMC will create TIFF files that conform to the specifications set forth in GPO's *FDsys Operational Specification for Converted Content (Version 3.3).*

GPO's requirement for access derivatives is that no single ocr-ed PDF file that is publicly available should exceed 10 MB. TIFF files created and delivered to GPO shall include tagging that allows GPO to create access derivatives that follow logical breaks within the publication in order to provide usable access derivatives.

In addition to the preservation-level files, the contractor or contractors may choose to provide GPO with derivatives of the content that may be optimized for other uses (e.g., screen or press optimized PDF files).

Each publication scanned and digitized, will have metadata associated with each TIFF file for preservation purposes as described above. During the planning process, Internet Archive will work with GPO to establish the manner in which GPO will provide descriptive metadata about each publication. The Archive will incorporate this metadata in MARC format together with all image data in the same container file which will provide encapsulation.

The Archive recognizes that the required metadata list is evolving and is subject to change over time. We will review the final list of metadata elements to be published by GPO before contract award, and will come to a mutually agreed-upon schema prior to the start of scanning operations.

12. Handling of Fragile, Rare and other materials not to be Dis-bound
The scanning workflow process of the Internet Archive includes procedures for handling fragile, rare and other materials not to be dis-bound. (add hyphen; may need global search and replace) The Scribe is designed for non-destructive scanning of bound materials and the manual operation process accommodates special handling for fragile materials.

Conversion Services will preserve the integrity of GPO source materials, capturing all physical and structural characteristics including the full layout of each individual page, the original sequence of pages within a given title and the original publication sequence of issues within a title.

The Internet Archive also has the capability to scan loose pages, foldouts and other oversized materials through alternate technologies that have been tested in-house and in commercial environments. The Archive's experience scanning rare and fragile books as well as items that cannot be dis-bound enables a comprehensive service offering to the GPO for its mass digitization needs.

The GPO states in the RFP that it may keep fragile and/or non-standard materials out of the digitization workflow, and also that the service provider may flag items to be kept out of the workflow and returned to the GPO or FDLP source library. Although the Archive believes that all necessary equipment and technologies are on hand for non-standard formats, we will work with the GPO to ensure that items are selected out at GPO's preference.

Task 4: Quality control of scanned file

There are four major phases to the QC process:

1) Before the materials are uploaded-At the book loading and scanning station; the scanners looks for; amongst other things, missing pages, crop/deskew problems, page marking (title page, front/back cover, tissue paper, first page of table of contents and notes any defects in the book (i.e. Missing/torn pages).

2) After the images are uploaded, derived and available via an URL- a statistical sampling and QA is conducted within the Scanning Center. Per ANSI z1.4 1993 Table 1, General Level 2.

3) Before the curation and bill is generated- An internal random audit is conducted outside the scanning center before the final curation approval and bill is generated.

4) After the materials are received by the GPO and the DS. Errors brought to IA's attention will be dealt with in a timely basis, within the specification of GPO's QC review outlined in the RFP (30 days for review, 10 days for correction, unless otherwise mutually agreed to in the planning session.) A decision will be made by IA as to whether it is best to rescan the material or fix it post-derive. The timeframe for the library to identify errors that will be fixed by the IA at no-charge shall be detailed in the digitization plan.

Scanning Center QA:

- IA uses ANSI z1.4 1993 Table 1, General Level 2 http://www.proqc.com/dl/aql.pdf
- Each day the scanning center will review a set of books from the previous days scanning. The number of books to QA depends on the total number of books in the set.

A. QA Process steps

Books are inspected for errors on-line, using the relevant files for each coded error type; found in the digital book record to look for errors or defects. Errors or defects, if found, are noted and added to the IA meta-manager form. An automatic scoring is then performed and a "pass/fail" grade is assigned to the lot.

Explanation- If 125 books were scanned in a period to be inspected, bin 5 would be selected. According to the truth table above, if there were 1 Major error or less and 2 Minor errors or less, the lot is passed. If there are 2 or more major defects or 3 or more Minor rejects, the lot fails. See below for what happens after this. The major/minor detail is show below in B5. Note: for major defect found during QA, they will be repaired on that book even if the lot passes.

- If the lot passes, the Scanning Center will approve all books (Curate).
- If a "fail' is generated, the Scanning Center Coordinator will review the errors/defects to ascertain if the errors were generated from outside the Scanning Center (for example a missing access file error would be sent to engineering for review) or from within the Scanning Center (for example a missing page).
- If the error was generated from within the Scanning Center the Coordinator would follow a pre-determined set of process steps ultimately culminating in a recommendation to deviate or approve the lot or a portion of the lot with appropriate corrective actions identified. At this stage the Book's Director or the Headquarters QA staff person is involved and must approve a deviation. A corrective action report will be generated for rejected lots. This will be reviewed with management for longer-term solutions or corrective action. This is done daily.

B Rescanning process-

For materials that are to be rescanned, a request to pull those books requiring rescanning is submitted to the Library; usually once a month. Materials are pulled, scanned and the original item is removed from the Internet Archive search engine and a new URL is assigned. This new URL is sent to the GPO and the DS along with the old URL for reference. A bug report could be the means to track this process.

C. Meta Manager, the post scanning reporting tool

This is the reporting tool that the Library may use to search and review books that have been scanned, uploaded, QA'd and then curated. The curation stage is the last stage in the IA process where the books are made viewable to the Library. This may happen on a non-scheduled basis but is typically done several times a month. Fields seen by GPO and DS in the Meta-Manager view will also be inspected, plus several internal fields. These fields will include:

- identifier
- title
- creator
- collection
- image count
- contributor
- sponsor
- sponsor date
- scandate
- curatenote
- curate date

D. Library card, required to view the Meta Manager

An IA library card and an email are required to view the metamanager (see steps listed below). Here is the process to access the meta data page.

1 Go to www.archive.org
2. Go to Patron info

3. Click on "get a virtual library card"
4. Follow instructions on page
5. Books may then be viewed that have been curated

Task 5: Delivery of digitized files back to GPO

The Internet Archive and/or LLMC, as appropriate, will deliver all tiff files generated as a result of the digitization process back to GPO on external storage devices within 10 days of acceptance by GPO, or as mutually agreed upon in the planning stage.

The Internet Archive proposes to provide tiff master files to the GPO on disk drives. An option for discussion during initial project planning would be the placement of a dedicated storage rack (with a full series of storage drives) at the Smithsonian Institution scanning center. As digital content is created, it can be stored directly onto these drives. As each drive is filled, it can be transported to the GPO.

Task 6: Delivery of original tangible materials back to GPO

Internet Archive Scanning Center Checkout Process
Scanning coordinator packs book into shipping cart/container per guidelines established between GPO and the Internet Archive. The operator creates and attaches the report communicating books rejected for scanning and identifying the rejection code.

The Internet Archive and/or LLMC will be responsible for shipping costs of returning original materials to the GPO, whether dis-bound or intact.

b. Equipment/Software that will be used for:

Scanning

The Internet Archive utilizes its open-source design Scribe scanner to convert bound monographs to digital format. This technology is being used at production scale and has been tested, as has the integrated image post-processing workflow, through large-scale digitization projects funded by the Alfred P. Sloan Foundation, Microsoft and several library and consortium partners.

The Scribe machine has a floating cradle that holds bound volumes at a fixed distance from two overhead digital cameras regardless of volume thickness. An angled glass platen is lowered on to the open volume to flatten the facing pages and produce the least distorted raw images possible. This minimizes the need for image curvature correction in the post processing steps, and results in a high quality, reproducible and printable image. An operator trained by the Internet Archive lowers the cradle after each two-page image is captured and turns pages manually. This has proven to be the fastest and least destructive process for scanning large quantities of bound materials. A picture of the Scribe in operation is included as Appendix C.

The 'standard' scanning center has 10 Scribes operating simultaneously for two shifts. The operation is scalable and more Scribes may be added to match capacity to demand. The Scribes operate much as the mini 'clusters' envisioned in the FDSys Conversion Guidelines, providing independent pipelines for scanned materials to the general collection. Moreover, the ability to utilize more than one scanning location for a large project ensures sufficient capacity and continued activity in the event of a service interruption at any single machine or scanning site.

- The Scribe scanner currently captures page images using a pair of digital single-lens-reflex (DSLR) cameras.

- The lighting system for illuminating the target books consists of eight (8) 5000 Kelvin, 36 degree, 35 watt museum-grade Solux bulbs, and provides a smooth daylight spectrum with a high color-rendering index.

- A lighting compensation program is used to help make the lighting even across images being scanned.

- Reference targets- a color target (ColorChecker 24) and a white card are shot with each book for reference, which can be used for ICC-based color management.

- Image transfer- Page images are downloaded in real time to a scanner management and image-processing computer that also runs the camera management software that releases the camera shutters.

- Equipment Calibration
    i. Cameras are calibrated per Manufacturer's specifications. Cameras out of spec or standard performance will be sent back to Manufacturer for repair.
    ii. Lights used in scanning process are replaced as is necessary. Light comp algorithms are run daily on each unit of scanning equipment.
    iii. Scribe stations are calibrated and aligned before being used.

OCR

The Internet Archive currently uses ABBYY FineReader OCR software to create XML files that contain the letter coordinates on the page. The files are then sandwiched with the JPEG-2000 image and packaged in a pdf file. Consequently the 'hidden' text can be highlighted, cut and pasted from the pdf document. Moreover, full-text search can be done on the pdf images resulting in highlighted terms within the text of the file. These coordinates are used when constructing PDF, DJVU (when produced), and Flipbooks to aid in highlighting terms on images of pages.

Documentation for the ABBYY FineReader software is available at http://www.abbyy.com/support/?param=2170.

Metadata

Each digitized publication will have associated with its TIFF and jpg2000 file(s) at least the minimal level of metadata for preservation purposes as identified in the GPO Metadata specification. The data elements will consist of bibliographic, technical, and administrative information necessary to track, manage, and preserve the associated files with each title for the future content management system. The TIFF data elements and values (e.g. presented in XML as fields with values associated with file header tags) represent metadata used to render and manage image data.

Bibliographic metadata will be stored in binary MARC, MARC-xml, and Dublin Core formats. The Internet Archive will request an up-to-date MARC record be provided for each of the titles to be digitized.

Technical metadata that describes the capture and QA components are captured in XML format. Structural metadata practices are evolving rapidly, with the Internet Archive drawing upon the work of Penn State University for automatically extracting structure in serials, for instance.

Descriptive metadata can be extracted from OCLC or Catalog of U.S. Government Publications if it is available through MARC records associated with each item to be digitized. The Internet Archive will work with GPO to identify the most expeditious way to obtain descriptive metadata for each item.

The RFP states that no single tiff file should be greater than 10 MB in size; if a bound volume results in a file greater than this limit, it will be split into smaller files with appropriate metadata added to enable reconstitution of complete volumes.

b. Process and means for delivery of digitized content back to GPO

The Internet Archive proposes to provide the digital tiff files to the GPO on disk drives. An option for discussion during initial project planning would be the placement of a dedicated storage rack (with a full series of storage drives) at the Library of Congress scanning center. As digital content is created, it can be stored directly onto these drives. As each drive is filled, it can be transported to the GPO.

c. Process and means for delivery of tangible content back to GPO

Scanning coordinator packs book into shipping cart/container per guidelines established between GPO and the Internet Archive. The operator creates and attaches the report communicating books rejected for scanning and identifying the rejection code.

The Internet Archive and/or LLMC will be responsible for shipping costs of returning original materials to the GPO, whether dis-bound or intact.

The Internet Archive does not request exclusivity; although the GPO may benefit from the consistent quality, file formatting and public access services provided by the Internet Archive.

2. Law Library Microform Consortium

a. Process for digitizing content:

Content is selected from among many print collections on offer to LLMC from its members based on its suitability for scanning and preservation. Most of the material is shipped to the main LLMC plant in Kaneohe where it is first validated for completeness. The texts are then scanned, via either high-speed scanners or with step-and-repeat scanners depending upon whether or not they can be dis-bound. Following scanning all images are individually proofed and tagged to the original pagination, and any imperfect images are immediately re-scanned. For new content, where metadata has not yet been created, copies of the tiffs are then sent to Saint Louis University Law Library, LLMC's cataloging partner, where the text are cataloged to highest OCLC standards and the records are added to OCLC's WorldCat Collections database for access by libraries interested in adding the records to their local electronic catalogs.

To accommodate situations where unique source documents for LLMC content are deemed too fragile or rare to ship to Kaneohe for scanning, the Consortium has created a special extern scanner program that enables onsite step-and repeat scanning. There are currently four such sites around the country; at the George Washington University Law Library, Saint Louis University Library, Los Angeles County Law Library, and the Hawaii State Archives. An agreement in principle for an additional site has been reached with the Library of Congress, with the final contractual details expected to be finalized within weeks and installation to come within months.

All images scanned by LLMC are delivered to our data service partner, National Business Systems, Inc (NBS), headquartered in Eagan, Minnesota. NBS is responsible for OCRing the LLMC images, storing the OCRed content, and hosting the web site that delivers the data to LLMC's hundreds of subscribing libraries. The data is delivered in PDF format using the latest version of Adobe Acrobat. The choice of Adobe Acrobat as the display engine was made to take advantage of the fact that, since all users have free and easy access to Adobe's system upgrades, LLMC's interface can develop naturally as Adobe itself improves.

LLMC's service partner, NBS, was established in 1972 as a data entry service company and subsequently expanded massively by offering service solutions associated with complete Document/Film Conversion Services, Print & Mail Services, Document Design & Reformatting, Personalized Direct Mail Marketing, Data Base Management, One To One Marketing Communications and Data Capture Services to a large number of national financial firms for which it provides a national web based Document Archive. Its Web Archives allow its customers to access the data in a centralized secure location. Its robust storage capacity provides a secure backup to LLMC own local storage, providing for the mirrored preservation of the full corpus of LLMC images.

b. Description of Quality Control process:

All scanned images, whether produced onsite or through LLMC's extern scanner program, are individually proofed by staff for quality and completeness. Unacceptable images are rescanned and all workflow is meticulously tracked and reported. Because of the heavy emphasis on preservation mandated by LLMC's principle customer base, academic law libraries, all of LLMC's online texts, are tracked on a publicly accessible database down to the page level. Subscribers use this online tool to determine whether they can safely dispose of their local hardcopy. This online tracking, done in conjunction with a non-profit organization called the Legal Information Preservation Alliance (LIPA), in effect recruits our entire subscriber base as "backup proofers" for the completeness of our product. To facilitate response from this backup network, a 'Feedback' mechanism has been incorporated in the LLMC-Digital interface so that any quality control issues discovered can be reported immediately and the problem resolved quickly.

As a supplement to tracking the completeness of the LLMC online product, LLMC and LIPA are also partnered in an effort to preserve in cooperating libraries a minimum number of copies of the hardcopy for all materials scanned. The locations of these backup preservation copies are also tracked using the publicly accessible database described above. LLMC itself has accepted responsibility for preserving one of the minimum numbers of copies retained. It fulfills that function by permanently storing all of the paper that comes into its purview through its high-speed scanning operations in a dark archive maintained in a commercial salt mine (make two words?) storage facility in Kansas.

In addition to image quality, LLMC also stresses the bibliographic integrity of its metadata. Part of its public commitment to its subscribers is to offer bibliographic data cataloged to the highest contemporary standards in OCLC's WorldCat Collections system. The work is done on a contract basis by Saint Louis University Law Library, an institution with a deservedly high technical reputation in this field and one of the few law libraries in North America qualified to do final-copy OCLC cataloging.

c. Equipment/Software that will be used for:

Scanning:

Currently LLMC does all of its step-and-repeat scanning (i.e. for bound materials) on special scanners made by a German company called SMA. This equipment has been chosen, both because it produces an exceptionally good image, but also because it has a very gentle impact on the materials being scanned. For materials that can be dis-bound, LLMC currently uses high-speed scanning equipment marketed by the Cannon Company.

OCR:

One of the reasons LLMC elected to partner with NBS, a company that services major financial corporations, is that we benefited by the significantly more robust technical infrastructure which that clientele demands and can pay for. NBS is currently utilizing ABBYY Recognitions Server 1.0. ABBYY is the most respected name in the industry as an OCR and PDF conversion solution. NBS plans to upgrade to Server 2.0 in the near future as this solution allows for PDF/A output that some of our customers may desire over the regular PDF output. The technical specifications can be found at http://www.abbyy.com/sdk/?param=53407. LLMC is confident both that current NBS OCR is done to the highest standards possible in the industry, and also that they will have access to improved technology in this field as soon as it occurs.

Metadata Capture:

In a move much lauded by its principal base, the academic law library community of North America, LLMC stresses the bibliographic integrity of its metadata. Part of its public commitment to its subscribers is bibliographic data cataloged to highest contemporary standards in OCLC WorldCat Collections system.

As part of this initiative, LLMC will always accept the GPO metadata. LLMC and Internet Archive will decide on a common metadata structure that we will both use to enable the files to be aggregated into a single collection. LLMC is experienced at integrating metadata from separate sources as its metadata operation is based at Saint Louis University Law Library under the expert direction of Richard C. Amelung, Ph.D., Professor of Legal Research Associate Director; serving on the Library of Congress' Working Group on the Future of Bibliographic Control committee; and Chairperson of the LLMC Board of Directors

IV. ADDITIONAL SERVICES

Hosting

While high resolution digital formats and multiple presentation formats will be provided to the GPO, the Internet Archive also proposes to permanently host the digitized documents for free permanent public access (PPA). The files will be stored in the Archive's Digital Repository along with over two petabytes of additional public collections.

The Digital Repository utilizes open source clustered storage technologies developed by the Internet Archive and refined over nearly 12 years of operational experience. The benefit of hosting GPO's files in this shared environment is that all the files will be subject to the same preservation policies, technology platform migrations and evolving standards that are applied to the Internet Archive's general collections. Savings resulting from scale and cumulative experience result in low operating costs, and the Archive's status as a 501(c) 3 not for profit organization ensures a low overhead burden. This low cost structure enables the Archive to offer very low all-in digitization costs.

Access

The Internet Archive is unique in its ability to provide free permanent public access (PPA) to the digitized collections via multiple widely used channels, including:

- Internet Archive website/government documents collections
- Public.resource.org
- OpenLibrary.org
- One Laptop Per Child/"Hundred Dollar Laptop" devices
- Print on demand optimized for the Espresso book printing machine

All GPO documents digitized by Internet Archive will be searchable via metadata, and in future full text searchable based on technologies being refined and deployed in the OpenLibrary.org project.

Documents scanned by Internet Archive will be automatically formatted for multiple presentation formats (derivative file formats) and print-on-demand options to provide for a wide spectrum of public access options.

V. WORK PLAN FOR DIGITIZATION

Months 0-3: Project Planning in Collaboration with the GPO

- Identify sourcing of materials for Phase I digitization
- Estimate total number of pages to be scanned in Phase I, and format types
- Plan for scanning location routing process – LoC, other IA center, LLMC
- Establish communication process & project key points of contact at GPO, LLMC and IA
- Establish process for midcourse assessments and/or adjustments to scanning process or pace

Months 4-6: Phase I Scanning for First Priority Documents Identified by GPO

- GPO to identify and transport the materials to Internet Archive and/or LLMC, pursuant to planning discussions
- The maximum digitization output rate from the Library of Congress scanning center as currently staffed is approximately 9-10 million pages per year. Space is available to double that capacity. Additional scanning centers of the Internet Archive and LLMC will provide sufficient overflow capacity to scan the entire legacy collection, if awarded, within 5 years.
- The OCR process is integrated in the scanning workflow and will keep pace with scanning activity.

Months 6-12: Phase II Scanning

Month 12: Program Scoping for Year 2

- IA and partners to estimate funds available for the project in year 2; provide forecast to GPO

Months 13-60

- Scanning to proceed on pace to complete the entire legacy collection, or awarded portion, by the end of the 5-year period.
- Annual meetings will be held with the GPO to provide forecasts of estimated funding and resulting scanning expectations

b. Process and means for delivery of digitized content back to GPO
As described above, the Internet Archive proposes to provide tiff master files to the GPO on disk drives. An option for discussion during initial project planning would be the placement of a dedicated storage rack (with a full series of storage drives) at the Smithsonian Institution scanning center. As digital content is created, it can be stored directly onto these drives. As each drive is filled, it can be transported to the GPO.

c. Process and means for delivery of tangible content back to GPO
The Internet Archive does not request exclusivity. There may be benefits to GPO, however, in having all documents scanned with the same quality, standards, optimized file formats and metadata through the Internet Archive's GovDocs program. This will

facilitate bulk or individual download, sharing and permanent public access to the entire GPO collection.

VI. SPECIFIC EXPERIENCE OF KEY PERSONNEL

Internet Archive Project Manager and Overall Project Director – Robert Miller
Robert Miller leads the Internet Archive's global book digitization project. In this capacity, he has three main roles; establishing and maintaining the relationships between the libraries and the funding partners, building and managing the teams that do the digitization and evangelizing within the library community to move more items from non-digital to digital. He built the team from scratch to its current level of 18 scanning centers in 5 countries.

Prior to the Archive, Robert co-founded 5 consumer product start up companies bringing over 85 products to market in the US, Europe and Australia. In addition he was CEO of FocusEngine, a venture capital funded search engine company. He has been featured in various publications such as the New York Times, WSJ, Inc and CNN.

Robert has two Fortune 500 company experiences; rising to sr. management roles in both Mattel Toys (consumer products) and AMP/Tyco (electronics).

Robert holds a BS in Industrial Engineering from Lehigh University in Pa.

Robert brings multi-cultural experience into the Archive; having lived in Afghanistan and Germany; and worked extensively in Asia and Europe.

Law Library Microform Consortium Project Manager: Kathleen Richman, Executive Director

FDLP Library Liaison: Ms. Judy Russell

Internet Archive Book Scanning Coordinator, Library of Congress - Ronnie Peoples
The Book Scanning coordinator in a Scribe center also carries the title of Site manager. Mr. Peoples' experience includes Account Management for records & file management at the Williams Lea Company and Processing Supervisor/section chief for the State Department doing document pre, imaging, image qa, data entry- 14 million passport records each year.

Before the State Department, Mr. Peoples was a Scanning Supervisor at the International Monetary Fund, responsible for document preparation, scanning and QA.

His overall responsibility now is to manage a professional, high quality 2 shift-scanning center capable of processing books at 500 pages per hour.

Overall responsibility for the scanning center rests with Mr. Peoples. He is charged with balancing quality scanning against output. He will work closely with his supervisor to be responsive to the needs of the Partner library; balanced against the goals and needs of IA. He has hiring/firing responsibilities contingent upon the guidelines set out by his supervisor, the Director of Books.

Specifics:

1. Interview, hire and train scanners capable of scanning at both a high level of quality and at a target rate of 500 pages/hr or above.
2. Create a positive work environment that balances individualism with teamwork.
3. Follow existing processes and procedures as outlined by engineering and operations management.
4. Maintain scanning equipment in accordance with preventative maintenance schedules.
5. Develop specific processes and procedures as is necessary to scan books at a consistent rate.
6. Establish and maintain good relations with the applicable library hosts.
7. Care for the books! Customize check in/checkout procedures for books in accordance with Partner Library.
8. Provide feedback to engineering team as is necessary on problems.
9. Provide clear and consistent communication to his management on issues that will impact cost, productivity, and quality or impact Partner Library relations.
10. At all times, be aware, that, he/she will be the main "face" of IA and as such will conduct himself and manage his team accordingly

Internet Archive Quality Control Manager, Library of Congress – Dorothy Gregory
Quality Control is the responsibility of the second site manager, in this case, Ms. Dorothy Gregory. She was previously the Preservation Imaging Specialist Supervisor at the National Archives and Records Administration (NARA,) and a Paralegal Specialist at the State Department. She began her career as an Automated Record Branch-Records Section Chief for imaging/microfilm at the State Department.

She shares the same site manager responsibilities as Mr. Peoples, with additional responsibility to conduct and oversee the QA process outlined above.

Internet Archive Scanning Specialists
There can be up to 20 scanning staff in each scanning center; rather than listing specific names the job description and responsibilities of each scanner are listed here.

Ability to create quality books scans while maintaining a rate of 500 pages per hour with the Scribe. This position directly impacts quality, cost and production. The Scanner reports to the Scanning Center Coordinator (SCC) or the Shift Supervisor.

Scanners are responsible for learning and mastering the Scribe program currently in use at IA scanning centers. Responsible for conducting QA on own work to ensure a quality scan is being provided to the IA and its contributors. Responsible for the care and handling of books as it pertains to the job duties. He/she will be able to recognize problems or anomalies within the workflow and inform the supervisor on duty for a resolution.

Skills required of all scanning staff, with training provided by the Internet Archive:

1. Learn the approved scanning method on the Scribe program and all of its functions.
2. Master all methods of Quality Assurance involved in the scanning process in order to assure the creation of a quality product.
3. Learn the various problems one can run across while working on collections in order to bring them to the attention of the supervisor.
4. Treat the books with utmost respect and care during the scanning process.
5. Pay strict attention to detail while scanning and remain focused on the job task at hand.
6. At all times be aware that he/she will be a "face" of IA and as such will conduct himself accordingly.
7. Adhere to all standards and policies as put forth in the Scanner hiring manual.

Metadata Specialist-
The scanning staff provides metadata services and quality control. The book loader creates the electronic record for each item to be scanned. The MARC record is pulled and the Dublin Core is created automatically. The scanner is responsible for reviewing the record for completeness.

OCR Manager- not required as OCR is done automatically, which increases the reliability.

VII. FINANCING STRATEGY

The Internet Archive is a not-for-profit organization that provides its services for cost recovery only. The Archive therefore has very limited ability to underwrite zero-revenue projects with its own operating funds. Funding for the Archive's digitization projects is usually secured before a project begins, either through fee-for-service arrangements with the library or libraries providing the materials or through third party grants.

The partners have outlined a strategy and have identified initial sources of funding to allow for scanning to begin immediately upon award of the contract. The short lead-time between publication and closing date of this RFP did not allow for identification of funding for the total legacy collection, estimated to cost $10-12 million. Moreover, since the GPO has reserved the right to earmark materials that may be scanned in-house or by other service providers, it is impossible to know precisely the total cost of the project. The structure of the project as proposed by the GPO therefore dictates an evolving funding strategy, raising funds as the project proceeds.

The Archive and its partners, the LLMC and Ms. Judith Russell, believe that sufficient funds can be raised over a five-year period to support the estimated $10-12 million cost to digitize the entire non-microform legacy collection. The partners have developed a financing strategy to pursue over the proposed five-year span of the project. The financing strategy involves three distinct sources: LLMC members for legal materials in the legacy collection, FDLP partner libraries and third party grants.

a. LLMC Members

Some of the 26% of the corpus for which LLMC will be responsible has already been scanned through funds provided by LLMC members' annual subscription fees. If LLMC is selected as a partner, it will donate this content as part of this project. To finance the remainder of the portion for which LLMC is responsible, LLMC and its partners will seek 3rd party grants to assist with its costs.

b. FDLP Partner Libraries

Over 1,250 libraries participate in the Federal Depository Library Program (FDLP) and commit their own staff and other resources to providing no-fee public access to a wide array of government publications in print and other tangible formats and through a growing array of online services. Public funds have been used to produce and disseminate this information and the 1,250 depository library collections are valuable public assets that support personal, professional and scholarly research. It is noteworthy that approximately 10% of the depository libraries preserve and provide access to print collections in excess of 1 million volumes.

Over 90 percent of new publications managed by and through the FDLP are available online, whether or not they are also available in tangible formats. This allows broad public access to and improved discovery of relevant content. However, most of the government information published between 1879 and 2000 is not available in electronic form for easy and effective no-fee public access.

The depository library community and the many users who rely on these libraries for public access to government information have been strong advocates for the systematic digitization of the legacy print collections to improve discovery and access as well as for the preservation of content. What the public wants and needs is 24/7 online access to a comprehensive collection of past, present and future government information, supported by the research services and tangible collections managed by the depository libraries.

The depository libraries have already demonstrated their commitment to public access to government information through their self-funded participation in the FDLP. There are 52 regional depository libraries that are required by law to retain their tangible collections, but all 1,250 depository libraries could more effectively manage their print collections and better service their constituencies if they had assurance of current and future access to the content in a manner that provides easy and effective no-fee public access. Many of them recognize the value of the digitization initiative proposed by the Government Printing Office (GPO) and will provide funding to support the initiative as well as providing access to their collections and other in-kind services to facilitate this effort. Some initial pledges have already been made and others will be confirmed by the Archive with Ms. Russell's assistance during the six months following award of the contract. These commitments will be far easier to secure once the contract has been awarded.

The funding commitments already secured from FDLP libraries will follow budget years, with either fixed dollar amounts or percentages of budget earmarked for this project. This per-annum flow of funding may be a factor in the pacing and total amount of scanning to be done by the Archive.

c. Third-Party Grants

Additional support will be sought from individuals and institutions outside the depository library community who recognize the value of establishing a comprehensive digital collection in a manner that both supports no-fee pubic access and permits aggregation and re-dissemination of the content through other initiatives.

We have identified several foundations to approach with specific proposals if the contract is awarded pursuant to this proposal. The great public good to be achieved by providing easy public access to the FDLP legacy collection has raised the level of funding interest. These organizations will only make commitments, however, after the contract has been awarded.

The Internet Archive's GovDocs program will be the starting point for pursuing third party funding, and may already encompass some of the legacy documents within the scope of the GPO RFP.

Internet Archive's GovDocs Program

The Internet Archive, in partnership with public.resource.org, has begun a digitization and access program for government documents called the GovDocs project. The Boston Public Library is the first Contributing Library in the program, and has agreed to lend a 50-year run of Congressional Hearings from 1936–1986, as well as a complete copy of the Catalog of Copyright Entries.

Phase 1 of the project has in its first few months produced approximately 2.5 million pages of digital text using a scanning and optical character recognition (OCR) technology suite developed by the Internet Archive. With existing funding the Archive expects to scan approximately 2.5 million additional pages. A screen image of the home page of this collection is included in Appendix D. Scanning is taking place at the Internet Archive's Northeast Regional Scanning Center located at the Boston Public Library.

Phase 1 is funded by a $250,000 matching grant challenge from Omidyar Network, which was met with a matching grant of $250,000 from the Kahle-Austin Foundation. Phase 2 of the program is envisioned to produce a full digital archive of government documents, consisting of the entire corpus of federal publications such as the Congressional Record and the Federal Register. Digitization of the GPO's legacy collection is a natural step into the second phase of this program, and in fact a portion of the collection to be digitized may already be available in the GovDocs collection.

The scanned documents produced by this project will remain in the public domain and be available for permanent public online access, download and sharing without restriction. In addition to becoming a part of the libraries at the Internet Archive and the Boston Public Library, the project will actively solicit and promote use of these data by commercial, noncommercial, and government groups. Images from this collection as they appear on the searchable website are included in Appendix D.

The Internet Archive sees the FDSys system and the GPO's plans for legacy collection digitization as a natural addition to this program. Through the GovDocs project the Archive has begun working with government documents librarians from several large FDLP libraries and can, in the provision of services to the GPO, apply its resulting understanding of the workflows, standards and culture of this set of partners.

The opportunity to join into a program that is underway with a community of government documents librarians is an expected benefit to the GPO in its digitization efforts. By working with the Internet Archive, GPO will tap into a growing movement of librarians with aligned interests, standards and access requirements. These librarians already act as stewards of government publications, and provide a mechanism for coordinating standards and requirements of the digitization process.

The Internet Archive proposes to meet with the GPO project staff periodically to discuss the forecasted amount of available funding and, consequently, the quantity of material that can be scanned in a given period.

This process is carries some uncertainty with it from year to year. However, we feel that the GPO's commitment to open access suggests that the digitization contract should be awarded to a fully open access service provider, not a private one. This means dependency on third party funding, an activity the Internet Archive has successfully pursued in the process of digitizing over 100 million pages of material.

APPENDIX B: SIMILAR EXPERIENCE MATRICES

a. FedLink contracts with Government agencies-

1. Dept of Interior- $17,000
2. Dept of Comptroller- $6,800
3. Dept of Treasury- $25,000
4. Smithsonian Institution - $200,000
5. Library of Congress- >$1.5 million

b. Non-government awards- (selected)

1. Biodiversity Heritage Library- $2.5 million
2. Boston Library Consortium- $1.0 million
3. Microsoft Networks (MSN)->$10.0 million
4. Omidyar/Kahle-Austin Foundations – GovDocs Collection- $500,000

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

1) ***Project Name/Contract Title and Contract Number- Biodiversity Heritage Library , (BHL) Scanning project***	
2) **Performed By:**	Internet Archive, (IA)
3) **Major Subcontractor(s):**	N/A- All work assigned to the Internet Archive was completed by the Internet Archive team.
4) **Key Personnel:**	Robert Miller, Director of Books, Internet Archive- Robert@archive.org ; Cathy Norton- Exec Director BLC, cnorton@mbl.edu
5) **Agency/Company:**	A consortium of 5 major libraries and a number of smaller institutions in two countries; this is a sub group of the Encyclopedia of Life- www.eol.org
6) **CO:**	n/a
7) **COTR:**	n/a
8) **Other Technical POC:**	n/a
9) **Period of Performance:**	2006-2008 and beyond
10) **Contract Type and Total Value:**	Appx $2.5 million dollars; 25+ million pages, initially. The project will continue past 2008 as future funding is obtained. A pilot phase was conducted and a two year funding stream was put in place.
11) **Product/Service Provided:**	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.
12) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.	
13) **Problems Encountered/Resolution**: No major problems encountered. Specs were developed, items were tested and signed off on, and production is currently ongoing in multiple locations in two countries. Specs were developed, items were tested and signed off on, and production has proceeded in two locations. Good communication channels have been put in place, dedicated people have been assigned to monitor the project and ensure quality throughput.	
14) **Awards, Recognitions, and Certifications Received:** Our contract has been expanded as new libraries have joined the project.	

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

15) ***Project Name/Contract Title and Contract Number- MSN Scanning project***

16) Performed By:	Internet Archive, (IA)
17) Major Subcontractor(s):	N/A- All work assigned to the Internet Archive was completed by the Internet Archive team.
18) Key Personnel:	Robert Miller, Director of Books Internet Archive- Robert@archive.org, Jay Girotto, Program Manager MSN, jgirotto@exchange.microsoft.com
19) Agency/Company:	MSN
20) CO:	n/a
21) COTR:	n/a
22) Other Technical POC:	n/a
23) Period of Performance:	2006-2008
24) Contract Type and Total Value:	>$10 million dollars; 100+ million pages
25) Product/Service Provided:	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.

26) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.

27) **Problems Encountered/Resolution:** No major problems encountered. Specs were developed, items were tested and signed off on, and production went on in 6 locations in two countries. Good communication channels were put in place, dedicated people were assigned to monitor the project and through put.

28) **Awards, Recognitions, and Certifications Received:** Our contract was renewed several times after being initially awarded.

29) Project Description and Approach: The goal was to digitize book and book-like items from 6 major libraries and return digital copies to both the funder, MSN, and to the libraries themselves, if they choose that option. A special portal was built by MSN and data digitized by the IA 'loaded' their data base. A back up copy was kept by the IA and was also used as a presentation copy for non-MSN viewing.

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

30) ***Project Name/Contract Title and Contract Number-Library of Congress, LC0 7D 7702***	
31) Performed By:	Internet Archive, (IA)
32) Major Subcontractor(s):	N/A- All work assigned to the Internet Archive was completed by the Internet Archive team.
33) Key Personnel:	Robert Miller, Director of Books Internet Archive- Robert@archive.org ,Mike Handy Program Manager, mhan@loc.gov
34) Agency/Company:	Government Printing Office
35) CO:	Vanessa Fox, vfox@loc.gov
36) COTR:	Anne Harrison, anha@loc.gov
37) Other Technical POC:	n/a
38) Period of Performance:	2007—2011
39) Contract Type and Total Value:	Indefinite Delivery Quantity Contract, Dec 4, 2007 to Sept 30, 2008, with 4 one year option periods. Value- > $1.5 million
40) Product/Service Provided:	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.
41) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.	
42) **Problems Encountered/Resolution:** As part of the normal steps involved in ramping up a high volume operation and blending LOC procedures with IA processes, new workflow guidelines have been and are being implemented as is necessary. Initial specs had been developed and items have been tested. Production is currently underway at the Library of Congress in a manner acceptable to both the IA and the funder. Production is managed by IA staff who are responsible for all elements listed above in item (11). Good communication channels have been put in place, and dedicated, trained people have been assigned to monitor the project and ensure quality throughput.	
43) Awards, Recognitions, and Certifications Received:	
44) **Project Description and Approach:** The goal is to digitize book and book-like items from the funder. A back up digital copy will be kept by the IA and will also be used as a presentation copy. Unlimited downloads are permitted. Information will be public access. The funder will retrieve the digital copies through the internet.	

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

45) ***Project Name/Contract Title and Contract Number-Department of the Treasury -LC07 D7702***	
46) Performed By:	Internet Archive, (IA)
47) Major Subcontractor(s):	N/A. All work assigned to the Internet Archive was completed by the Internet Archive team.
48) Key Personnel:	Robert Miller, Director of Books Internet Archive- Robert@archive.org Judy Lim-Sharpe who can be reached at 202-622-0990, Judy.LIM-SHARPE@do.treas.gov
49) Agency/Company:	Department of the Treasury,
50) CO:	Vanessa Fox, vfox@loc.gov
51) COTR:	Anne Harrison, anha@loc.gov
52) Other Technical POC:	n/a
53) Period of Performance:	2008—2009
54) Contract Type and Total Value:	➢ $25,000 USD, 250,000+ pages
55) Product/Service Provided:	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.
56) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.	
57) **Problems Encountered/Resolution:** No major problems have yet been encountered. Specs were developed, items have been tested and signed off on. Production will be starting shortly at the IA scanning center located at the Library of Congress. Good communication channels have been put in place, dedicated and well-trained people have assigned to monitor the project and throughput.	
58) **Awards, Recognitions, and Certifications Received:**	
59) **Project Description and Approach:** The goal was to digitize book and book-like items from the Funder and return digital copies to the funder. A back up copy will be kept by the IA and will also be used as a presentation copy for public access viewing.	

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

60) *Project Name/Contract Title and Contract Number-Department of Interior, LC0 7D 7702*

61) Performed By:	Internet Archive, (IA)
62) Major Subcontractor(s):	N/A. All work assigned to the Internet Archive was completed by the Internet Archive team.
63) Key Personnel:	Robert Miller, Director of Books Internet Archive- Robert@archive.org, George D. Franchiois who can be reached at 202-208-3796,. George_D_Franchois@nbc.gov.
64) Agency/Company:	Department of Interior
65) CO:	Vanessa Fox, vfox@loc.gov
66) COTR:	Anne Harrison, anha@loc.gov
67) Other Technical POC:	n/a
68) Period of Performance:	2008—2009
69) Contract Type and Total Value:	$ 17,000 dollars; 170,000 pages
70) Product/Service Provided:	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.

71) **Small Business Participation: Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.**

72) **Problems Encountered/Resolution:** No problems have yet been encountered. Specs have been developed and items were tested and signed off on. Production will be in the IA managed scanning center located at the Library of Congress. Good communication channels have been put in place, dedicated and trained people have been assigned to monitor the project and ensure quality throughput.

73) **Awards, Recognitions, and Certifications Received:**

74) **Project Description and Approach:** The goal is to digitize book and book-like items from the Department of the Interior. A back up digital copy will be kept by the IA and will also be used as a presentation copy. Unlimited downloads are permitted. Information will be publicly accessible. The funder will retrieve the digital copies through the internet.

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

75) ***Project Name/Contract Title and Contract Number-Comptroller of the Currency -LC07 D7702***

76) Performed By:	Internet Archive, (IA)
77) Major Subcontractor(s):	N/A- All work assigned to the Internet Archive was completed by the Internet Archive team.
78) Key Personnel:	Robert Miller, Director of Books Internet Archive- Robert@archive.org, Sabrina Pacifici who can be reached at 202-874-4722, Sabrina.Pacifici@occ.treas.gov
79) Agency/Company:	Comptroller of the Currency
80) CO:	Vanessa Fox, vfox@loc.gov
81) COTR:	Anne Harrison, anha@loc.gov
82) Other Technical POC:	n/a
83) Period of Performance:	2008—2009
84) Contract Type and Total Value:	➢ $6,852.70 USD, 68,750+ pages
85) Product/Service Provided:	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.

86) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.

87) **Problems Encountered/Resolution:** No major problems have yet been encountered. Specs were developed; items have been tested and signed off on. Production will be starting shortly at the IA scanning center located at the Library of Congress. Good communication channels have been put in place, dedicated and well-trained people have assigned to monitor the project and throughput.

88) **Awards, Recognitions, and Certifications Received:**

89) **Project Description and Approach:** The goal was to digitize book and book-like items from the Funder and return digital copies to the Funder. A back up copy will be kept by the IA and will also be used as a presentation copy for public access viewing.

ATTACHMENT 1
SIMILAR EXPERIENCE MATRIX TEMPLATE

90) ***Project Name/Contract Title and Contract Number- Boston Library Consortium Library, (BLC) Scanning project***	
91) **Performed By:**	Internet Archive, (IA)
92) **Major Subcontractor(s):**	N/A- All work assigned to the Internet Archive was completed by the Internet Archive team.
93) **Key Personnel:**	Robert Miller, Director of Books, Internet Archive- Robert@archive.org ; Barbara Preece- Exec Director BLC, bpreece@blc.org
94) **Agency/Company:**	A consortium of 19 major Libraries in New England. http://www.blc.org/library_user_services/mem_cats.html
95) **CO:**	n/a
96) **COTR:**	n/a
97) **Other Technical POC:**	n/a
98) **Period of Performance:**	2007-2008 and beyond
99) **Contract Type and Total Value:**	Approx $1 million dollars; 10+ million pages, initially. The project will continue as funding is obtained in future years. A pilot was conducted and then a two-year funding stream was put in place.
100) **Product/Service Provided:**	Document handling, meta data collection (MARC 21 and Dublin Core), image capture (color), storage file generation, access file creation, assignment of persistent identifier, OCR generation, 100% QA, document return, lifetime management of files, unlimited downloads. Special handling or features as discussed.

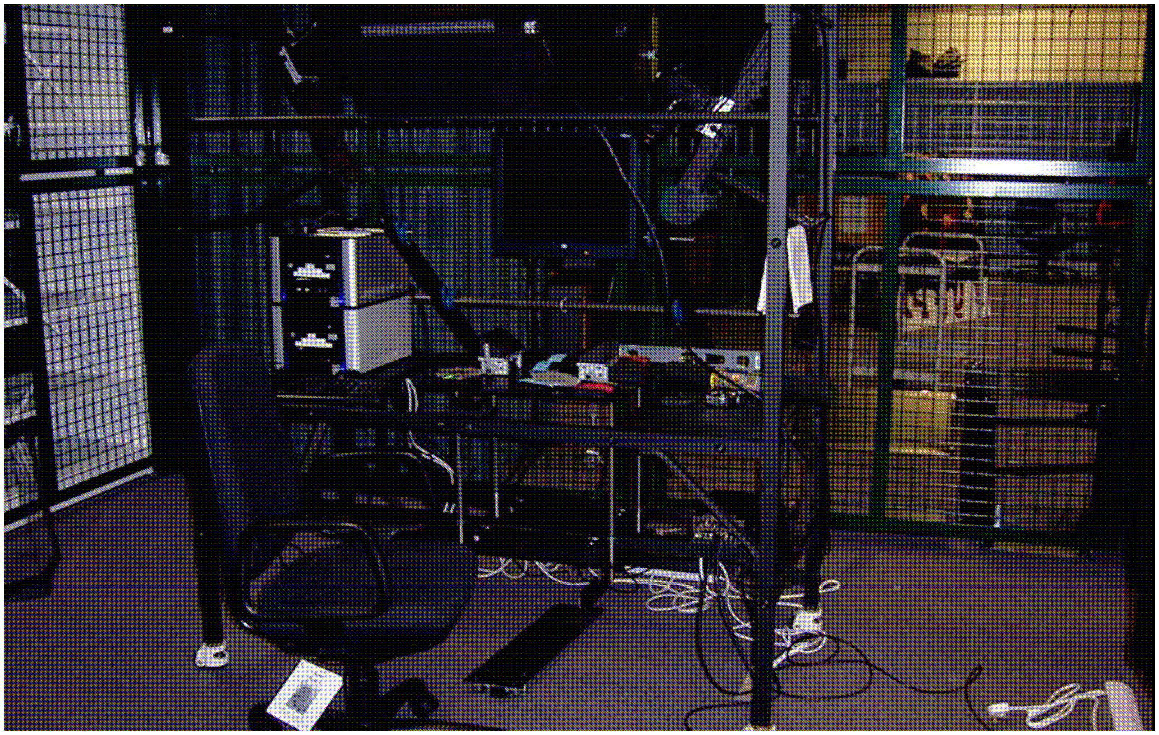
101) **Small Business Participation:** Internet Archive is a 501 (c) (3) , with business offices located in California, USA, Canada and Europe.

102) Problems Encountered/Resolution: No major problems encountered. Specs were developed, items were tested and signed off on, and production is ongoing in Boston Ma, serving 19 libraries who ship materials to a central scanning center. Specs were developed, items were tested and signed off on, and production has proceeded.. Good communication channels have been put in place, dedicated people have been assigned to monitor the project and ensure quality throughput**.**

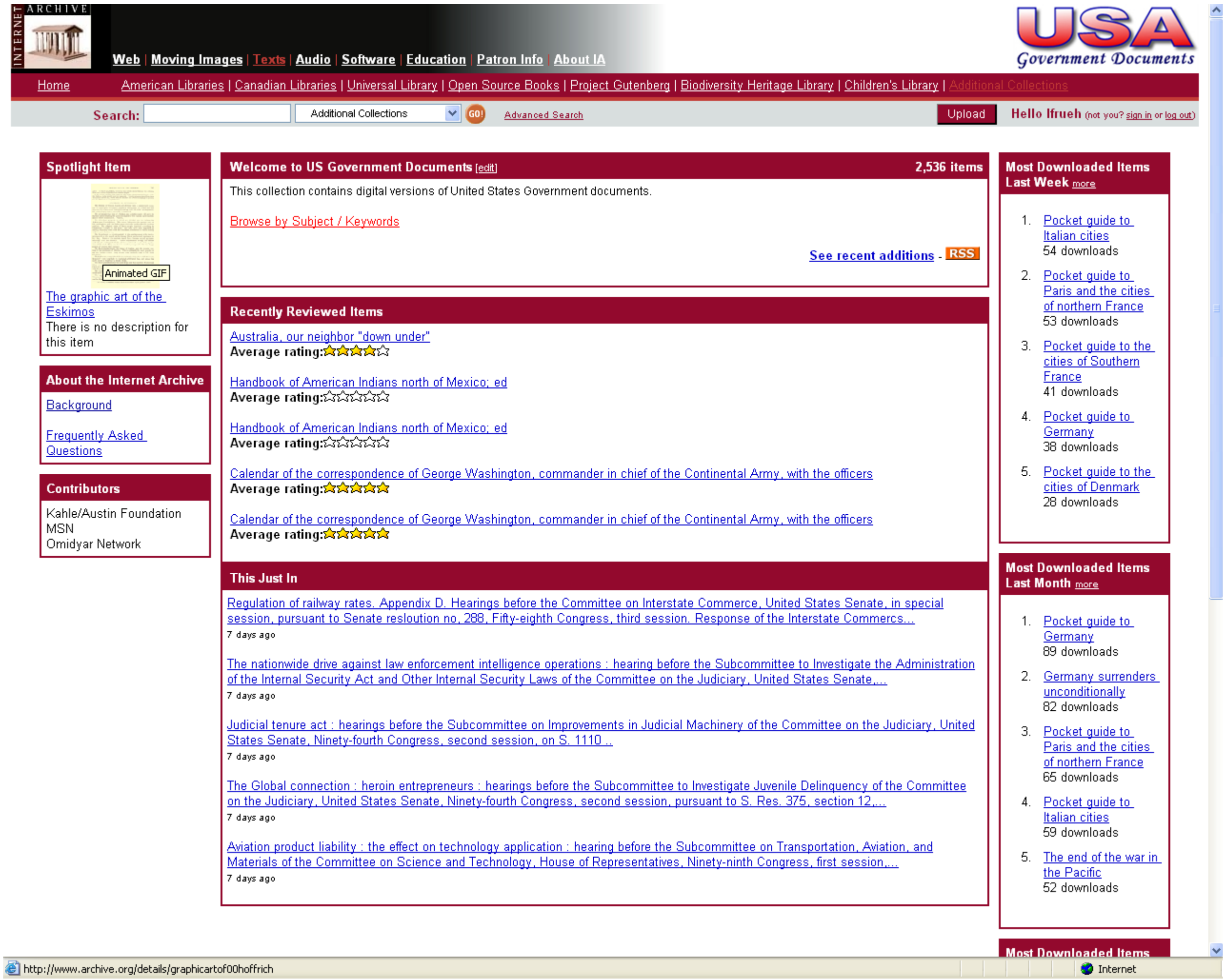
103) **Awards, Recognitions, and Certifications Received:** Our contract has been expanded as new libraries have joined the project.

104) **Project Description and Approach:** The goal was to digitize book and book-like items from 19 major state, public and private libraries and return digital copies to the libraries themselves, if they choose that option. BLC users with existing library web sites are pointing their users back to digitized copies stored and served by IA. Unlimited downloads are permissible.

APPENDIX C: SCRIBE SCANNER



APPENDIX D: IMAGES FROM INTERNET ARCHIVE'S GOVDOCS COLLECTION


INTERNET ARCHIVE
Web | Moving Images | Texts | Audio | Software | Education | Patron Info | About IA
USA Government Documents
Home
American Libraries | Canadian Libraries | Universal Library | Open Source Books | Project Gutenberg | Biodiversity Heritage Library | Children's Library | Additional Collections
Search:
Additional Collections
GO!
Advanced Search
Upload
Hello lfrueh (not you? sign in or log out)
Animated GIF
http://www.archive.org/details/graphicartof00hoffrich
Internet

Spotlight Item

The graphic art of the Eskimos
There is no description for this item

About the Internet Archive

Background

Frequently Asked Questions

Contributors

Kahle/Austin Foundation
MSN
Omidyar Network

Welcome to US Government Documents [edit] **2,536 items**

This collection contains digital versions of United States Government documents.

Browse by Subject / Keywords

See recent additions - RSS

Recently Reviewed Items

Australia, our neighbor "down under"
Average rating:★★★★☆

Handbook of American Indians north of Mexico; ed
Average rating:☆☆☆☆☆

Handbook of American Indians north of Mexico; ed
Average rating:☆☆☆☆☆

Calendar of the correspondence of George Washington, commander in chief of the Continental Army, with the officers
Average rating:★★★★★

Calendar of the correspondence of George Washington, commander in chief of the Continental Army, with the officers
Average rating:★★★★★

This Just In

Regulation of railway rates. Appendix D. Hearings before the Committee on Interstate Commerce, United States Senate, in special session, pursuant to Senate resloution no. 288, Fifty-eighth Congress, third session. Response of the Interstate Commercs...
7 days ago

The nationwide drive against law enforcement intelligence operations : hearing before the Subcommittee to Investigate the Administration of the Internal Security Act and Other Internal Security Laws of the Committee on the Judiciary, United States Senate,...
7 days ago

Judicial tenure act : hearings before the Subcommittee on Improvements in Judicial Machinery of the Committee on the Judiciary, United States Senate, Ninety-fourth Congress, second session, on S. 1110 ..
7 days ago

The Global connection : heroin entrepreneurs : hearings before the Subcommittee to Investigate Juvenile Delinquency of the Committee on the Judiciary, United States Senate, Ninety-fourth Congress, second session, pursuant to S. Res. 375, section 12,...
7 days ago

Aviation product liability : the effect on technology application : hearing before the Subcommittee on Transportation, Aviation, and Materials of the Committee on Science and Technology, House of Representatives, Ninety-ninth Congress, first session,...
7 days ago

Most Downloaded Items Last Week more

1. Pocket guide to Italian cities
54 downloads
2. Pocket guide to Paris and the cities of northern France
53 downloads
3. Pocket guide to the cities of Southern France
41 downloads
4. Pocket guide to Germany
38 downloads
5. Pocket guide to the cities of Denmark
28 downloads

Most Downloaded Items Last Month more

1. Pocket guide to Germany
89 downloads
2. Germany surrenders unconditionally
82 downloads
3. Pocket guide to Paris and the cities of northern France
65 downloads
4. Pocket guide to Italian cities
59 downloads
5. The end of the war in the Pacific
52 downloads

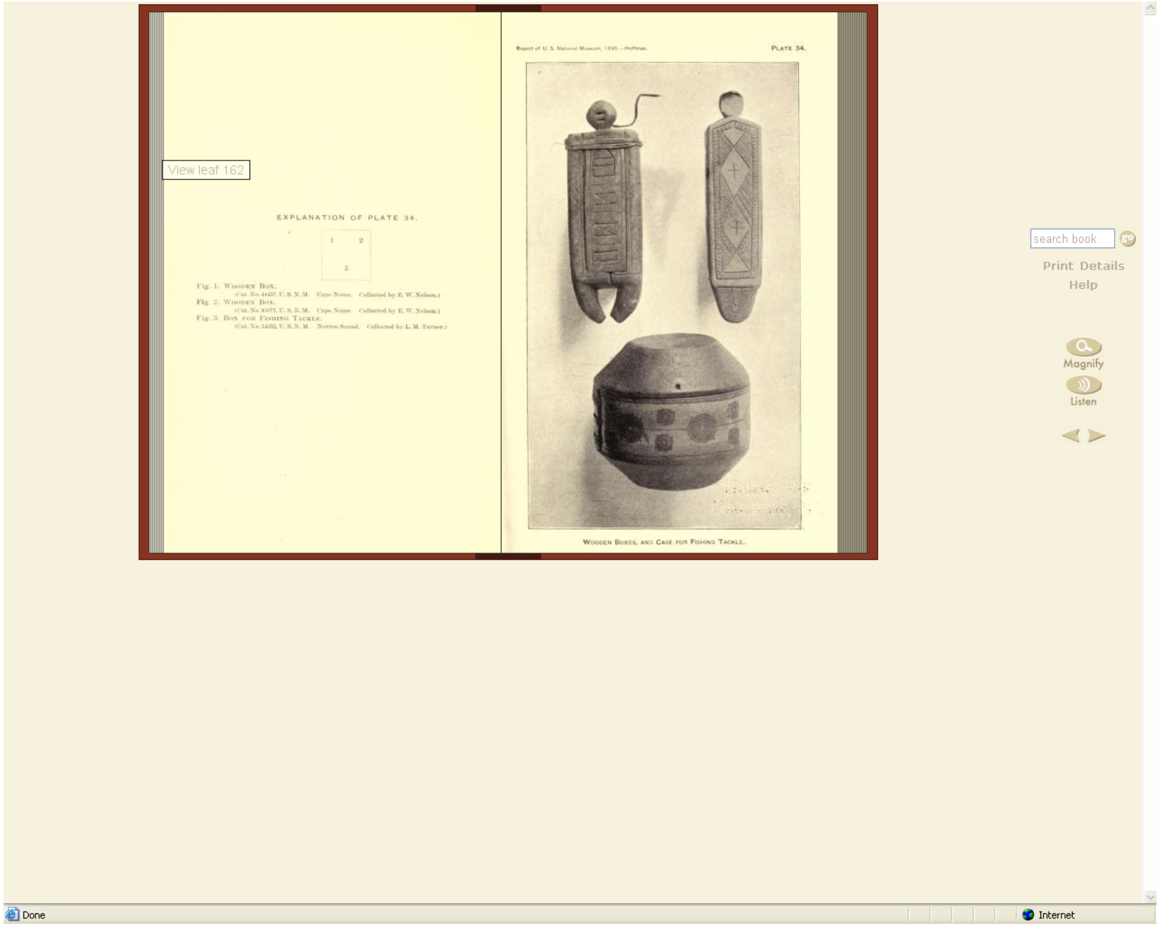
Most Downloaded Items

Sample Digitized Materials from Internet Archive Government Documents Collection (published by Government Printing Office)

FlipBook Interface:

http://www.archive.org/details/graphicartof00hoffrich

The graphic art of the Eskimos (1897)



EXPLANATION OF PLATE 34.

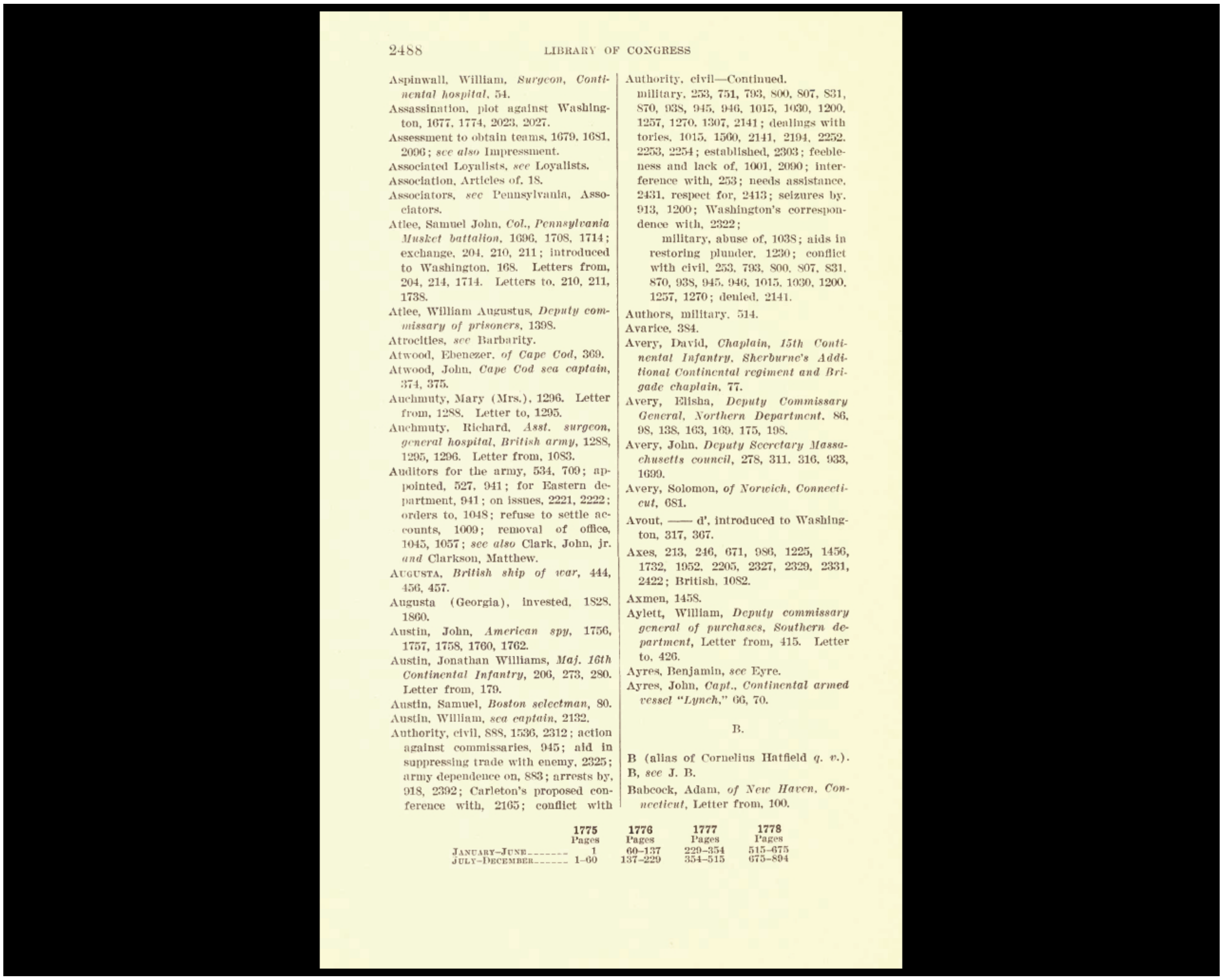
Fig. 1. Wooden Box.
(Cat. No. 44457, U. S. N. M. Cape Nome. Collected by E. W. Nelson.)
Fig. 2. Wooden Box.
(Cat. No. 33977, U. S. N. M. Cape Nome. Collected by E. W. Nelson.)
Fig. 3. Box for Fishing Tackle.
(Cat. No. 24592, U. S. N. M. Norton Sound. Collected by L. M. Turner.)

Report of U. S. National Museum, 1895.—Hoffman. Plate 34.

Wooden Boxes, and Case for Fishing Tackle.

PDF Format:

Calendar of the correspondence of George Washington, commander in chief of the Continental Army, with the officers (1915)



Aspinwall, William, *Surgeon, Continental hospital*, 54.

Assassination, plot against Washington, 1677, 1774, 2023, 2027.

Assessment to obtain teams, 1679, 1681, 2096; *see also* Impressment.

Associated Loyalists, *see* Loyalists.

Association, Articles of, 18.

Associators, *see* Pennsylvania, Associators.

Atlee, Samuel John, *Col., Pennsylvania Musket battalion*, 1696, 1708, 1714; exchange, 204, 210, 211; introduced to Washington, 168. Letters from, 204, 214, 1714. Letters to, 210, 211, 1738.

Atlee, William Augustus, *Deputy commissary of prisoners*, 1398.

Atrocities, *see* Barbarity.

Atwood, Ebenezer, *of Cape Cod*, 369.

Atwood, John, *Cape Cod sea captain*, 374, 375.

Auchmuty, Mary (Mrs.), 1296. Letter from, 1288. Letter to, 1295.

Auchmuty, Richard, *Asst. surgeon, general hospital, British army*, 1288, 1295, 1296. Letter from, 1083.

Auditors for the army, 534, 709; appointed, 527, 941; for Eastern department, 941; on issues, 2221, 2222; orders to, 1048; refuse to settle accounts, 1009; removal of office, 1045, 1057; *see also* Clark, John, jr. *and* Clarkson, Matthew.

AUGUSTA, *British ship of war*, 444, 456, 457.

Augusta (Georgia), invested, 1828, 1860.

Austin, John, *American spy*, 1756, 1757, 1758, 1760, 1762.

Austin, Jonathan Williams, *Maj. 16th Continental Infantry*, 206, 273, 280. Letter from, 179.

Austin, Samuel, *Boston selectman*, 80.

Austin, William, *sea captain*, 2132.

Authority, civil, 888, 1536, 2312; action against commissaries, 945; aid in suppressing trade with enemy, 2325; army dependence on, 883; arrests by, 918, 2392; Carleton's proposed conference with, 2165; conflict with

Authority, civil—Continued.

military, 253, 751, 793, 800, 807, 831, 870, 938, 945, 946, 1015, 1030, 1200, 1257, 1270, 1307, 2141; dealings with tories, 1015, 1560, 2141, 2194, 2252, 2253, 2254; established, 2303; feebleness and lack of, 1001, 2090; interference with, 253; needs assistance, 2431, respect for, 2413; seizures by, 913, 1200; Washington's correspondence with, 2322;

military, abuse of, 1038; aids in restoring plunder, 1230; conflict with civil, 253, 793, 800, 807, 831, 870, 938, 945, 946, 1015, 1030, 1200, 1257, 1270; denied, 2141.

Authors, military, 514.

Avarice, 384.

Avery, David, *Chaplain, 15th Continental Infantry, Sherburne's Additional Continental regiment and Brigade chaplain*, 77.

Avery, Elisha, *Deputy Commissary General, Northern Department*, 86, 98, 138, 163, 169, 175, 198.

Avery, John, *Deputy Secretary Massachusetts council*, 278, 311, 316, 933, 1699.

Avery, Solomon, *of Norwich, Connecticut*, 681.

Avout, —— d', introduced to Washington, 317, 367.

Axes, 213, 246, 671, 986, 1225, 1456, 1732, 1952, 2205, 2327, 2329, 2331, 2422; British, 1082.

Axmen, 1458.

Aylett, William, *Deputy commissary general of purchases, Southern department*, Letter from, 415. Letter to, 426.

Ayres, Benjamin, *see* Eyre.

Ayres, John, *Capt., Continental armed vessel "Lynch,"* 66, 70.

B.

B (alias of Cornelius Hatfield *q. v.*).

B, *see* J. B.

Babcock, Adam, *of New Haven, Connecticut*, Letter from, 100.

	1775 Pages	**1776** Pages	**1777** Pages	**1778** Pages
JANUARY–JUNE	1	60–137	229–354	515–675
JULY–DECEMBER	1–60	137–229	354–515	675–894

APPENDIX E: RESUMES OF KEY PERSONNEL